

Обучение машинного перевода без параллельных текстов*

Строганов А. А. Бахтеев¹ О. Ю. Стрижов² В. В.

¹Московский физико-технический институт

²Вычислительный центр им. А. А. Дородницына ФИЦ ИУ РАН

Данная работа посвящена исследованию машинного перевода без использования параллельных текстов. Исследование сконцентрировано на использовании нейросети с несколькими моделями Seq2seq для перевода с одного языка на другой и обратно. Особенностью данных моделей является то, что исходный текст кодируется во внутреннее представление модели, а затем декодируется в текст на другом языке. Две модели Seq2seq имеют общее скрытое пространство. Примером, иллюстрирующим работоспособность данного алгоритма, будет использован эксперимент по взаимному переводу с двух похожих языков – русского и украинского.

Ключевые слова: *машинный перевод, нейросеть, Seq2seq.*

1 Введение

Благодаря недавним достижениям в области глубокого обучения и наличию крупномасштабных параллельных корпусов, машинный перевод достиг впечатляющей производительности на нескольких языковых парах. Тем не менее, эти модели работают очень хорошо, только если они снабжены огромным количеством параллельных данных в порядке миллионов параллельных предложений. К сожалению, параллельные корпуса стоят дорого [2], поскольку они требуют специализированного опыта и часто не существуют для языков с низким уровнем ресурсов.

Есть несколько подходов к построению оптимального метода обучения. Предлагается использовать рекуррентные нейронные сети с короткой и долгой памятью и нейронные сети, в которых реализовано внимание. В других методах используются нейронные сети, которые осуществляют перевод в два этапа. Такой метода называется Seq2Seq [3].

Данная работа посвящена последнему методу последовательного перевода. Предлагается с помощью первой рекуррентной нейронной сети, основанной на долгой памяти перевести входящую последовательность в вектор, а с помощью второй перевести этот вектор в выходную последовательность на нужном нам языке [1]. Данный метод позволяет гораздо быстрее обучить нейронную сеть переводу с одного языка на другой, в связи с использованием ей предыдущего опыта и наличию у нее памяти и внимания. Проверка и анализ метода проводятся с помощью алгоритма BLEU (Bilingual evaluation understudy) для проверки качества текста, переведенного с одного языка на другой на паре языков русский-украинский.

2 Постановка задачи

Во время обучения нет параллельных пар предложений. Предполагаем, что нам подойдет модель, отображающая предложения из обоих языков в одно общее векторное пространство.

*Работа выполнена при финансовой поддержке РФФИ, проект №00-00-00000. Научный руководитель: Стрижов В. В. Авторы: А. В. Грабовой, О. Ю. Бахтеев, В. В. Стрижов, Eric Gaussier, координатор Малиновский Г. С. Консультант: Бахтеев О. Ю.

Мы будем использовать модель, в которой используются главных юнита: encoder f и decoder g . f и g в нашей конкретной модели - две рекуррентные нейронные сети. Задача f - отображать предложения в латентное пространство(сразу для двух языков) и g - отображать из латентного пространства в предложения(первого языка и второго соответственно). Введем обозначения: $D^{src} = [s_1^{src}, \dots, s_{m_{src}}^{src}]$, $D^{tgt} = [s_1^{tgt}, \dots, s_{m_{tgt}}^{tgt}]$.

Для реализации этого метода определим функционалы, которые будут минимизироваться. Чтобы модель не обучилась возвращать в конце цикла исходные данные, необходимо зашумить исходные предложения. Пусть $\sigma(x)$ - результат наложения шума на слово x . Оптимизировать будем следующую функцию:

$$L_{AE} = ||d(e(\sigma(x))) - x||^2$$

Пусть дана какая-то модель слабого перевода \hat{g} . На втором шаге функция потерь будет иметь вид:

$$L_{TR} = ||d(e(\hat{g}(e(x)))) - x||^2$$

Пусть дана модель D , различающая скрытые представления векторов предложений из двух языков. На последнем шаге оптимизируем дискримантор, чтобы он различал представления векторов разных языков в скрытом пространстве:

$$L_{ADV} = \log p(lang = src | Encoder(x)) + \log p(lang = tgt | Encoder(y))$$

В итоге нужно минимизировать следующую функцию:

$$L = a * L_{AE} + b * L_{TR} + c * L_{ADV} \longrightarrow \min$$

здесь a, b, c - калибруемые гиперпараметры.

Литература

- [1] Kyunghyun Cho, Bart Van Merriënboer, Dzmitry Bahdanau, and Yoshua Bengio. On the properties of neural machine translation: Encoder-decoder approaches. *arXiv preprint arXiv:1409.1259*, 2014.
- [2] Philipp Koehn. *Statistical machine translation*. Cambridge University Press, 2009.
- [3] Ron J Weiss, Jan Chorowski, Navdeep Jaitly, Yonghui Wu, and Zhifeng Chen. Sequence-to-sequence models can directly translate foreign speech. *arXiv preprint arXiv:1703.08581*, 2017.

Поступила в редакцию